KazuNet: Invariant Learning for Factor Graphs

Yueze Wang
University of California Santa Barbara
yueze@ucsb.edu

Abstract

Nonlinear factor graphs combine heterogeneous residual scales, poorly conditioned coordinates, and gauge freedoms. KazuNet learns a shared update rule in an observable quotient space. A four-stage pipeline calibrates factor scales, whitens the observable curvature, predicts invariant weights and curvature coefficients, and selects a finite update using two frozen merit functions. The 45,294-parameter network uses Weyl–Procesi scalars and force/curvature Gram features; a matrix-free implementation supports larger states. Strict Mode preserves observable $\mathrm{GL}(n)\times\mathcal{G}_{\mathrm{scale}}^{\mathrm{struct}}\times S_K$ *equivariance, while Practical Mode adds diagonal-equivariant spectral truncation. Across eight real or standard benchmarks with 1,371 cases and 19 solvers, KazuNet wins 83.0%–90.8% of non-tied paired comparisons against five external neural baselines at five outer steps. Three controlled ablations further characterize scale robustness, coordinate conditioning, and sparse memory use. Line search requires no additional Jacobians, and sparse clique curvature reduces factor memory by* $508.8\times$ *at state dimension 2,436.*

1. Introduction

Nonlinear factor-graph optimization underpins bundle adjustment [2, 30], pose-graph SLAM [9, 16], non-rigid surface registration [28], and engineering state estimation [39]. Given state $\theta \in \mathbb{R}^n$ (or a local manifold chart [1]) and K factors (F_k, g_k, H_k), solvers minimize $E(\theta) = \sum_{k=1}^{K} \omega_k \rho_k(F_k(\theta))$ [6, 8, 15, 17, 20, 33] over fixed outer-step and exact Jacobian-evaluation budgets $T = J \in \{3, 5, 10, 20\}$.

Undamped, full-rank Gauss–Newton is algebraically equivariant under invertible coordinate transformations, and modern solvers such as Ceres [3, 23] include Jacobian column scaling. Yet practical damped solvers, robust M-estimators, normal-equation solves in finite precision, and raw-coordinate learned descriptors all depend on residual units and coordinate conditioning. Thus, changing the numerical chart or factor scaling can alter finite optimization trajectories even when the underlying physical estimation problem is identical. KazuNet removes this numerical coupling by calibrating factor scales and constructing observable quotient coordinates before predicting an update (Figs. 2 and 3).

Learning in quotient geometry. KazuNet separates geometric normalization from learned preconditioning (Fig. 2). Its contributions are:

1. **Calibrated quotient coordinates:** equilibrated rank estimation, structured factor-scale calibration, Foster bridge treatment, and square-root whitening map heterogeneous graphs into a common observable space.
2. **Invariant learned curvature:** a shared 45,294-parameter network predicts factor weights and conflict, damping, and secant coefficients from Weyl–Procesi invariants; sparse clique operators extend the same construction to larger states.
3. **Finite updates across tasks:** dual-merit backtracking converts the learned direction into an accepted step without extra Jacobians, improving five-step recovery across bundle adjustment, surface registration, and power-grid state estimation (Fig. 1).

2. Related Work

Classical nonlinear least squares and robust estimation. Large-scale geometric reconstruction and state estimation rely on Gauss–Newton, Levenberg–Marquardt [23], Dog-leg [24], manifold retractions [1], and problem-specific factorizations such as Schur-complement and variable-projection bundle adjustment [2, 18, 30]. Practical toolkits like Ceres [3] mitigate ill-conditioning through diagonal trust metrics and column scaling, while outlier resistance relies on robust M-estimators and IRLS [6, 17, 20, 35] or graduated non-convexity (GNC) [8, 15, 33]. Multi-sensor variance balancing follows Helmert variance-component estimation (HVCE) [11, 14, 27], whose single-step linear variance estimates can lose stability on under-redundant bridges or heavy contamination.

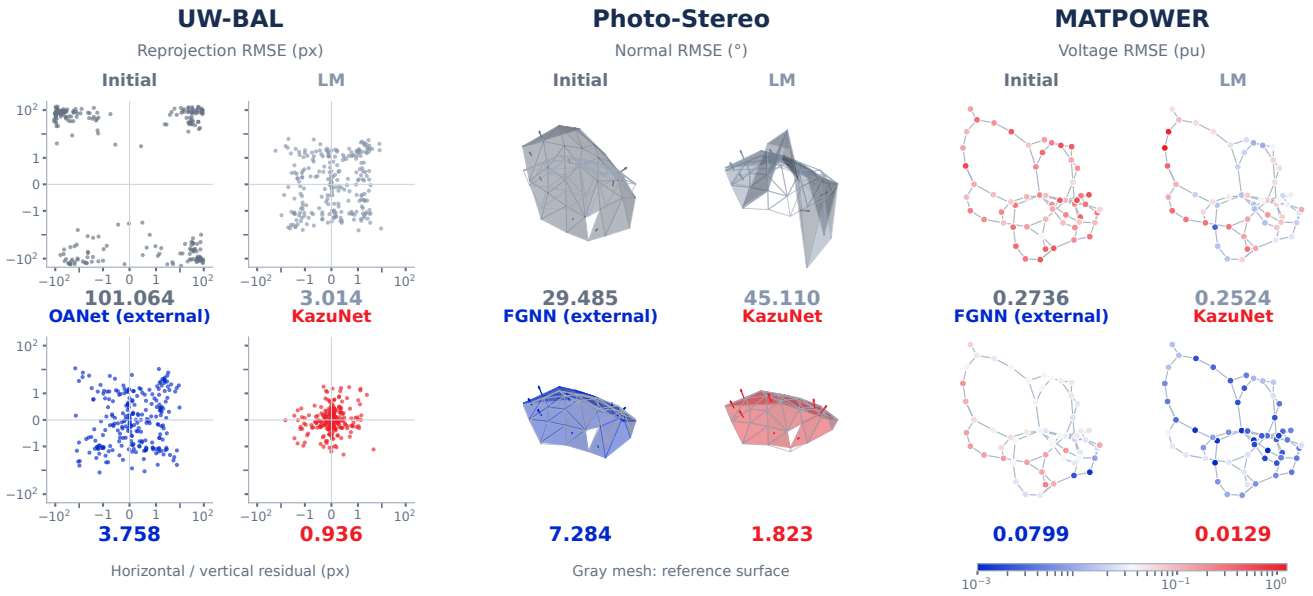


Figure 1. **Five-step recovery across three tasks.** Each task shows Initial and LM above, its best external backbone and KazuNet below. OANet and FGNN are selected by the full-task five-step mean. Labels report task errors. UW-BAL plots share symmetric-log residual axes (linear within ± 1 px); gray meshes mark reference surfaces; MATPOWER colors share a logarithmic per-bus voltage-error scale (pu).

Learned optimization and factor-graph networks. Learning-to-optimize methods began with coordinate-wise recurrent updates [4]. In geometric vision, DeepLM [19] embeds LM inside deep learning frameworks via automatic sparse Jacobians and stochastic domain decomposition, scaling least-squares computation rather than learning cross-task quotient preconditioning. Learned geometric correspondence and robust estimation networks—including LFGC-DFE [34], CN-IRLS [26], OANet [36], CLNet [38], NCMNet [22], and Factor Graph Neural Networks (FGNN) [37]—predict factor weights or updates directly from Euclidean descriptors. Alongside our in-repo PointNet-LM and Euclidean-Net baselines, such raw-coordinate predictors remain sensitive to $\mathrm{GL}(n)$ reparameterizations and arbitrary residual scaling.

Equivariant invariants and numerical preconditioning. Classical invariant theory characterizes polynomial group invariants via bilinear and trace contractions [25, 32], forming universal scalar parameterizations for equivariant networks [31]. Meanwhile, numerical linear algebra provides communication-optimal TSQR [13] and Jacobi-preconditioned Lanczos/PCG [7, 21]. KazuNet unites both: Stage 1–2 reduce $\mathrm{GL}(n) \times \mathcal{G}_{\mathrm{scale}}^{\mathrm{struct}}$ gauge transforms to an $O(r_{\mathrm{obs}})$ action via calibrated square-root whitening, allowing a compact invariant Transformer to predict anisotropic curvature and finite updates across domains.

3. Gauge-Quotient Formulation & Method

Let $\theta \in \mathbb{R}^n$ be the optimization state (or local manifold tangent step via $\mathrm{Retract}(\theta, \Delta\theta)$ [1]) governed by K active factors ($v_k \in \{0,1\}$, DOFs $m_k \geq 1$, nominal rank q_k^{nom}):

$$E(\theta) = \sum_{k:v_k=1} \omega_k F_k(\theta), \qquad g_k = \nabla_\theta F_k, \quad H_k \succeq 0, \quad H_k \approx \nabla_\theta^2 F_k. \tag{1}$$

where oracles optionally supply square-root Jacobians $J_{\sqrt{k}} \in \mathbb{R}^{m_k \times n}$ ($H_k = J_{\sqrt{k}}^\top J_{\sqrt{k}}$) for $n \leq 256$, or sparse clique blocks and Hessian-vector products $v \mapsto \sum_k w_k H_k v$ for $n > 256$. Two physical symmetry groups act on (1): (1) coordinate gauge transforms $\theta(\phi) = \theta_0 + T\phi$ ($T \in \mathrm{GL}(n)$) inducing $g_k \mapsto T^\top g_k$, $H_k \mapsto T^\top H_k T$, $J_{\sqrt{k}} \mapsto J_{\sqrt{k}} T$; and (2) a structured factor-scale group $\mathcal{G}_{\mathrm{scale}}^{\mathrm{struct}} = \mathbb{R}_{>0}^{|\mathcal{K}_{\mathrm{over}}|} \times \mathbb{R}_{>0}^{\mathrm{min}}$, scaling each overdetermined factor individually ($\omega_k > 0$, $s_{\mathrm{redundancy},k} > 0.05$) and the low-redundancy pool jointly ($\omega_{\mathrm{min}} > 0$, $s_{\mathrm{redundancy},k} \leq 0.05$), equaling full $\mathcal{G}_{\mathrm{scale}} = \mathbb{R}_{>0}^K$ whenever $\mathcal{K}_{\mathrm{min}} = \emptyset$, alongside permutation S_K. An update rule is **quotient-equivariant** if $\Delta\hat{\phi}$ and $\Delta\hat{\theta}$ induce identical observable changes in $\mathrm{range}(B_\theta)$.

A check denotes calibrated quantities, $(\check{F}_k, \check{g}_k, \check{H}_k) = \pi_k(F_k, g_k, H_k)$; a bar denotes their coordinates in the whitened basis W. The analytical precision π_k adjusts the factor scale, while the learned weight w_k subsequently adjusts its contribution to the update.

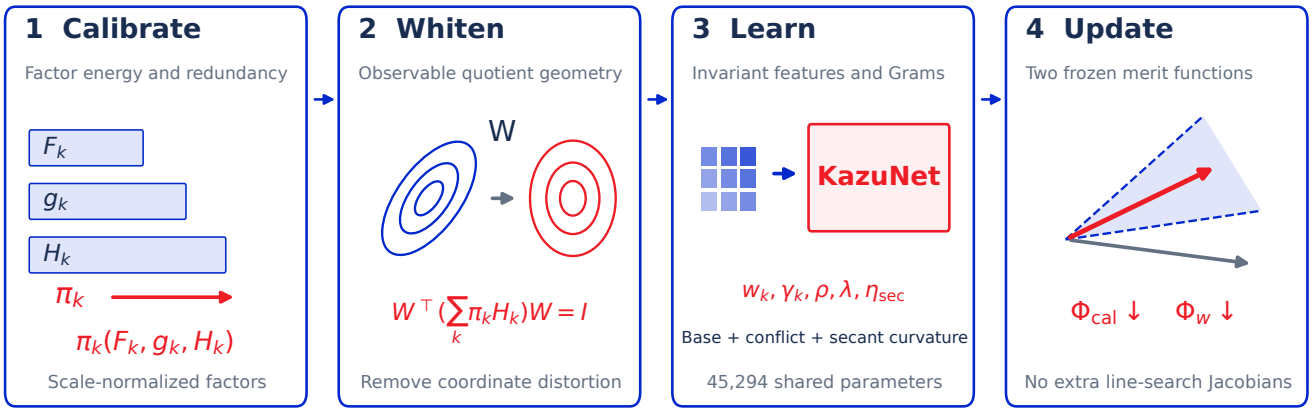


Figure 2. **KazuNet architecture.** Calibration aligns factor scales; whitening constructs observable coordinates; a shared network predicts invariant factor weights and curvature coefficients; the resulting direction is projected and tested against two frozen merit functions. Blue denotes geometric computation and red denotes learned coefficients and the accepted update.

3.1. Stage 1: Equilibrated Rank & $\mathcal{G}_{\text{scale}}^{\text{struct}}$ Calibration

Stage 1 estimates effective rank q_k and excess residual dimension $\nu_k = \max(m_k - q_k, 0)$ in equilibrated coordinates. It uses the row-Gram matrix $(J_{\sqrt{k}}W_0)(J_{\sqrt{k}}W_0)^\top$ when square-root Jacobians are available, and Jacobi-equilibrated curvature otherwise. The factor energy separates into tangent and orthogonal components:

$$e_k^{\parallel} = g_k^\top H_k^+ g_k, \qquad e_k^{\perp} = \max\left(2F_k - e_k^{\parallel},\, 0\right). \quad (2)$$

In the local quadratic model, $e_k^{\parallel}$ is the energy explained by tangent directions, and $e_k^{\perp}$ is the remaining residual energy. This separation connects precision estimation to the factor's residual redundancy. A smooth redundancy score $s_{\text{redundancy},k}$ separates overdetermined factors from a low-redundancy pool. Overdetermined factors use residual-based precisions with James–Stein shrinkage; pooled factors share a quantile-based precision. Full-graph Foster leverage identifies bridge factors, whose precision is refined through the minimal-pool projection. Supp. Sec. E retains the complete expressions and constants.

Calibration then executes **two quotient-Helmert passes** [11, 14, 27] with cross-pool quantile alignment $c_{\min}$, computing leverages $\ell_{0,k} = \text{tr}(W_p^\top(\pi_{\text{pass},k}H_k)W_p)$ and post-step normalized energies:

$$R_{\text{post},k} = 2\pi_{0,k}F_k + 2\pi_{0,k}g_k^\top d_p + \pi_{0,k}d_p^\top H_k d_p,$$
$$\hat{E}_k^{\text{raw}} = \frac{\max(R_{\text{post},k}, 0)}{\max(m_k - \ell_{0,k},\, 0.1m_k)}. \quad (3)$$

Here $\pi_{0,k}$ is the preliminary precision, W_p whitens the curvature under the current pass weights, and d_p is the corresponding local quadratic step. The residual energy is then divided by the unused degrees of freedom to refine the precision. We use $\hat{E}_{\text{med}} = Q_{0.30}(\{\hat{E}_k\})$, $\text{ex}_k = \max(\hat{E}_k/(3.5\hat{E}_{\text{med}} + 10^{-30}) - 1, 0)$, and Cauchy-Helmert weight $u_{\text{pre},k} = (1 + \text{ex}_k + 0.75\text{ex}_k^2)^{-1}$, producing $(\check{F}_k, \check{g}_k, \check{H}_k) = \pi_k(F_k, g_k, H_k)$.

Theorem 1 ($\mathcal{G}_{\text{scale}}^{\text{struct}}$-Invariance & Upper-Tail Breakdown)**.** *Assuming $(T, J_{\sqrt{k}})$ preserves non-degenerate numerical rank ($\kappa_2(J_{\sqrt{k}}T) \ll 10^{11}$), calibrated tuples $(\check{F}_k, \check{g}_k, \check{H}_k)$ are strictly invariant under $\mathcal{G}_{\text{scale}}^{\text{struct}} = \mathbb{R}_{>0}^{|\mathcal{K}_{\text{over}}|} \times \mathbb{R}_{>0}^{\min}$ ($\pi_k \mapsto \omega_k^{-1}\pi_k$, reducing to full $\mathcal{G}_{\text{scale}} = \mathbb{R}_{>0}^K$ when $\mathcal{K}_{\min} = \emptyset$), as $\omega_{\min}$ cancels identically in $\pi_0^{\min}$ and cross-pool alignment $c_{\min}$. For any outlier set $\mathcal{O}$ ($F_k \to +\infty$) with $|\mathcal{O}| < (1-q)K$, $Q_q(\{\hat{E}_k\}) \le B_{\text{in}} < \infty$, conferring $\approx 70\%/65\%$ upper-tail immunity (lower-tail boundary & proof in Supp. A.1).*

3.2. Stage 2: Singular Quotient Whitening

From total calibrated curvature $\check{B} = \sum_{k:v_k=1} \check{H}_k \succeq 0$ (or square-root stack $L_{\text{stack}} = [\sqrt{\pi_k}J_{\sqrt{k}}]_{k:v_k=1}$), Stage 2 constructs an observable quotient basis $W \in \mathbb{R}^{n \times r_{\text{obs}}}$ discarding modes below relative threshold $\tau_{\text{obs}} = 10^{-12}$:

$$B_{\text{eq}} = D\check{B}D = V\Lambda V^\top,$$
$$W = DV_{:,\text{obs}}\Lambda_{\text{obs}}^{-1/2}, \quad (4)$$
$$W^\top \check{B} W = I_{r_{\text{obs}}}.$$

where $D = \text{diag}(d_i^{-1})$ provides column equilibration. Whitening makes the aggregate curvature the identity on the observable subspace. Directions in its nullspace do not change the local measurements, so the learned update is computed only in the remaining coordinates.

Definition 1 (Observable Quotient vs. Primal Pullback)**.** *For rank-$r \le n$ curvature B_θ, transformed whitener W_ϕ under $\theta(\phi) = \theta_0 + T\phi$ satisfies $TW_\phi = W_\theta Q + N_\perp C$ for some orthogonal gauge $Q \in O(r)$ and nullspace matrix $N_\perp$ ($B_\theta N_\perp = 0$).*

Theorem 2 (Quotient Whitening Equivariance)**.** *Let $\Delta\hat{\theta} = W_\theta\xi_\theta$ and $\Delta\hat{\phi} = W_\phi\xi_\phi$ where $\xi(\{\bar{g}_k, \bar{S}_k\})$ is any*

orthogonal-equivariant solver of whitened contractions $\bar{g}_k = W^\top \check{g}_k$, $\bar{S}_k = W^\top \check{H}_k W$. *Then* $\xi_\phi = Q^\top \xi_\theta$, *yielding:*

$$T\Delta\hat{\phi} = \Delta\hat{\theta} + N_\perp (CQ^\top \xi_\theta). \tag{5}$$

When $r = n$, $N_\perp = 0$ *and exact primal equivariance holds* ($T\Delta\hat{\phi} = \Delta\hat{\theta}$). *When* $r < n$, *the nullspace shift lies strictly in* $\ker(B_\theta)$, *guaranteeing exact observable quotient equivalence:* $B_\theta(T\Delta\hat{\phi} - \Delta\hat{\theta}) = 0$.

Theorem 3 (TSQR+SVD Square-Root Stabilization)**.** *For* $n \le 256$ *with* $J_{\sqrt{k}}$, *computing* W *via column-equilibrated chunked TSQR followed by SVD [13] on* L_{stack} *preserves* $\sigma_i > 10^{-12}\sigma_{\max}$ *without forming normal equations, which square the condition number of the full-rank observable square-root system, maintaining* 4.86*e*-11 *relative step equivariance at* $\kappa_2(T) = 10^6$ *(Supp. Table S1).*

3.3. Stage 3: Weyl-Procesi Preconditioner & Risk

In the r_{obs}-dimensional quotient space, KazuNet consumes ten invariant node scalars, force and curvature Gram matrices, and an eight-dimensional spectrum summary. The scalars describe leverage, redundancy, calibrated energy, gradient magnitude, alignment, and factor role. A three-layer Transformer with width 32, four attention heads, and feedforward width 72 predicts factor weights w_k, conflict gains γ_k, isotropic mixing ρ, relative damping λ_{rel}, and a secant gain. Its 45,294 parameters require 4.35M MACs per step at $K = 64$, independent of state dimension. The force Gram matrix describes agreement between factor gradients, and the curvature Gram matrix describes overlap between their constrained directions. Both are unchanged by a common orthogonal change of quotient coordinates. Exact features, output ranges, and KazuNet-Zero settings are in Supp. Sec. E.

With $c_{\text{lev}} = r_{\text{obs}}^{-1} \sum_k w_k \ell_k$, Godambe conflict $\Omega_\xi \succeq 0$, and rank-1 secant curvature $\Omega_{\text{sec}} \succeq 0$, Stage 3 solves:

$$\begin{aligned} \bar{M}_0 &= (1-\rho)\bar{A} + (\rho + \lambda_{\text{rel}}) c_{\text{lev}} I_{\text{obs}}, \\ \bar{M} &= \bar{M}_0 + \Omega_\xi + \Omega_{\text{sec}} \succ 0, \\ \bar{M}\bar{d}_{\text{conf}} &= -\bar{g}_w. \end{aligned} \tag{6}$$

Here $\bar{A} = \sum_k w_k \bar{S}_k$ and $\bar{g}_w = \sum_k w_k \bar{g}_k$. The conflict term limits the curvature contributed by disagreeing factors:

$$\Omega_\xi = \tau_{\text{cap}}(\tau_{\text{cap}} I + \widetilde{\Omega})^{-1}\widetilde{\Omega}, \qquad \widetilde{\Omega} = \frac{\sum_k a_k \xi_k \xi_k^\top}{1 + 2\mu_{\text{bulk}}}.$$

The post-step force is $\xi_k = \bar{g}_k + \bar{S}_k \bar{d}_0$, where $\bar{d}_0 = -\bar{M}_0^{-1}\bar{g}_w$; $a_k \ge 0$ weights its uncertainty, μ_{bulk} is the median eigenvalue of the numerator, and $\tau_{\text{cap}} > 0$ bounds the conflict spectrum. The secant term adds curvature along the observed gradient change:

$$\Omega_{\text{sec}} = \frac{\chi_{\text{sec}}\eta_{\text{sec}}}{[h^\top z]_+ + \|z\|_2^2 + 10^{-12}} zz^\top.$$

Here $h = W^\top \check{B}\Delta\theta_{t-1}$, $z = W^\top \sum_k w_k \pi_k (g_{k,t} - g_{k,t-1})$, and χ_{sec} selects positive secant curvature; the term is zero without history. Supp. Sec. A.4 gives the scalar coefficients and numerical conventions.

For $n > 256$, a matrix-free Jacobi-preconditioned Lanczos/PCG branch [7, 21] replaces dense $O(n^3)$ eigensolves with Loewner-Woodbury conflict updates.

Theorem 4 (Anisotropic Spectral Risk & Non-Collapse Floor)**.** *For whitened Gaussian recovery* $y = \bar{J}x_\star + \varepsilon$ ($\text{Var}(\varepsilon_i) = \sigma^2$, *singular values* σ_i) *under Tikhonov [29] damping* $\lambda_i \ge 0$, *expected squared error splits into shrinkage bias plus noise variance:*

$$\begin{aligned} \mathcal{R}(\lambda) &= \sum_{i=1}^{r} \frac{\lambda_i^2 x_{\star,i}^2 + \sigma^2\sigma_i^2}{(\sigma_i^2 + \lambda_i)^2}, \\ \lambda_i^\star &= \frac{\sigma^2}{x_{\star,i}^2}. \end{aligned} \tag{7}$$

Whenever $\sigma_i^2 x_{\star,i}^2/\sigma^2$ *varies across modes, anisotropic* $\lambda_i^\star$ *strictly beats any constant isotropic* $\bar{\lambda}$. Non-collapse condition: *over unrestricted* $\lambda_i \ge 0$, $\lambda_i \to \infty$ *collapses* $\hat{x} \to 0$ *with* $\mathcal{R} \to \|x_\star\|_2^2$, *so higher isotropic floors hold under non-collapse progress constraints (the convex-duality calculation in Supp. Sec. A.3).*

3.4. Stage 4: Dual-Merit Descent & Symmetry Modes

Stage 4 maps $\bar{d}_{\text{conf}}$ through $\bar{M}$-oblique dual-cone projection $\Delta\hat{\theta} = W\mathcal{P}_{\text{cone}}(\bar{d}_{\text{conf}})$ certifying simultaneous descent on frozen anchors $\Phi_{\text{cal}}^{(t)}(\theta) = \sum_k \pi_k^{(t)} F_k(\theta)$ and $\Phi_w^{(t)}(\theta) = \sum_k w_k^{(t)}\pi_k^{(t)} F_k(\theta)$.

The projection has the one-line form

$$\mathcal{P}_{\text{cone}}(d) = d - \frac{[g^\top d + \eta ab]_+}{a^2}\bar{M}^{-1}g,$$

where $d = -\bar{M}^{-1}\bar{g}_w$, $g = \bar{g}_{\text{cal}}$, $a = \|g\|_{\bar{M}^{-1}}$, $b = \|d\|_{\bar{M}}$, $c = -g^\top d/(ab)$, and $\eta = \min\{0.02, (1-c^2)/2\}$. This expression applies to nonzero, noncollinear directions; Supp. Sec. A.4 specifies the degenerate cases.

The two merit functions measure calibrated factor energy and its learned reweighting. Their coefficients are held fixed while backtracking tests candidate step lengths, then recomputed at the next outer step. Candidate evaluation uses factor energies without additional Jacobians. The search includes straight and quotient-curved directions; Practical Mode also includes Jacobi-equilibrated spectral truncation with thresholds $\tau \in \{10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}\}$.

Proposition 1 (Strict and Practical Symmetries)**.** *Under the rank assumptions of Theorems 1–3, Strict Mode disables spectral truncation and preserves observable* $\text{GL}(n)\times$

$\mathcal{G}^{\text{struct}}_{\text{scale}} \times S_K$ *equivariance. Practical Mode adds Jacobi-equilibrated spectral truncation and preserves* $\text{Diag}_{>0}(n) \times \mathcal{G}^{\text{struct}}_{\text{scale}} \times S_K$ *equivariance; general orthogonal equivariance need not hold. Supp. Sec. A.3 gives the numerical examples.*

4. Experimental Protocol and Baselines

Training Objective and Evaluation Budgets. All learned solvers train identically on 15,360 synthetic multi-regime graphs in double precision (`float64`) on NVIDIA A100 GPUs, strictly disjoint from the $N = 256$ validation set and all evaluation suites, using unrolled $T = 5$ BPTT:

$$\mathcal{L} = 0.65\psi_{\text{H}}(e_T) + 0.20\log(e_T + 0.03) + \frac{0.15}{T-1}\sum_{t<T}\psi_{\text{H}}(e_t) + \mathcal{R}_{\text{reg}}. \tag{8}$$

where $\psi_{\text{H}}(u) = \sqrt{u^2 + 0.04} - 0.2$ and normalized step-t error e_t propagates gradients through Stage-3 MLP/attention heads, the Godambe-Woodbury linear solve $\bar{M}^{-1}\bar{g}_w$, curved updates, and θ_{t+1}, while $(\pi_k, W, f_k, A_{\text{force}}, A_{\text{curv}})$ and discrete line-search selection $j^\star$ are detached via stop-gradient at each outer step. One shared checkpoint per learned solver is evaluated across all tasks.

Budgets and baselines. Here T denotes outer steps and J denotes full Jacobian/curvature evaluations. We evaluate equal budgets $T = J \in \{3, 5, 10, 20\}$ and compare five-step KazuNet variants against 20-step baselines in the budget study; all methods start from the same initial states. The comparison includes six classical methods (LM [23], Dogleg [24], IRLS-Cauchy [17], BAR [6], GNC-GM [8, 15, 33], and HVCE [27]); three in-repo learned factor-graph baselines (CN-IRLS [26], Euclidean-Net, and PointNet-LM); five external neural backbones (OANet [36], FGNN [37], CLNet [38], NCMNet [22], and LFGC-DFE [34]); and four variants of our method (KazuNet-Zero, KazuNet, KazuNet-M, and KazuNet-L). LM and Dogleg use Ceres-style column scaling [3]. Because OANet, CLNet, NCMNet, and LFGC-DFE originate in two-view correspondence pruning (and FGNN in general factor-graph message passing), all five external backbones receive an identical 18-dimensional Euclidean factor descriptor and predict per-factor IRLS-LM weights trained end-to-end under Eq. (8). DeepLM [19] adds a nineteenth solver at the five-step budget, using the author implementation with the same task residuals and initial states; Supp. Sec. C details its native stopping rules and results. Model sizes and implementation details appear in Supp. Secs. B and D.

Eight Real/Standard Benchmarks (1,371 Cases) & Three Theoretical Ablations (600 Cases). We organize evaluation into public observations, real geometry and standardized infrastructure benchmarks, and controlled ablations: **Tier I (Unaltered Real Public Observations, 531 cases):** Raw-BAL [2] (140 cases; raw UW BAL poses and SIFT correspondences evaluated on 64 held-out tracks, px); Raw-g2o [10] (160 cases: 80 SE(3) + 80 SE(2) unanchored pose graphs with native information matrices, ATE m); and HPatches-SIFT [5] (231 cases: 115 six-frame and 116 three-frame SIFT symmetric-transfer homographies, px). **Tier II (Real Scans, SfM Windows, and MATPOWER Grids, 840 cases):** UW-BAL [18, 35] (150 post-RANSAC Schur/VarPro bundle adjustment windows, px); Wild-BAL (150 pre-RANSAC UW BAL scenes with $10^6\times$ matcher-scale spread, 20%–30% gross outliers, and corrupted rotation priors, px); 3D-Scans [12] (180 Stanford/AIM@SHAPE laser-scanned meshes under ARAP deformation, 22% occlusion, and flying-pixel outliers, mm); Photo-Stereo (180 uncalibrated photometric stereo scans with two-DOF height/albedo gauge, deg); and MATPOWER (IEEE 14/30/57-bus transmission grids [39], 180 cases with SCADA/PMU noise and bad data, pu). **Controlled Theoretical & Systems Ablations (A1–A3, 600 cases +** $C = 30$ **gauge grids):** A1: Scale-1e6 (200 cases, $10^6\times$ heterogeneous multi-sensor SLAM scaling, m); A2: Gauge-GL ($C = 6$ ARAP [120, mm], $C = 6$ BAL [120, px], $C = 12$ Dim. [240], plus $C = 30$ conditioning grids isolating Thms. 1–2 and Prop. 1); and A3: Sparse-Scale (160 sparse-clique meshes with $n \leq 2,436$ isolating matrix-free Lanczos-PCG scaling, mm).

Table 1. **Held-out validation ($N = 256$, $T = J = 5$).** Classical LM, in-repo and external learned baselines, and KazuNet variants. Bold and underline mark the best and second-best displayed values; component ablations appear in Sec. 5 and Supp. Sec. B.

Method	Params	Mean	Median
LM	0	1.9922	1.5388
PointNet-LM	45,389	0.5263	0.3770
CN-IRLS	44,611	0.5039	0.3372
OANet	55,477	0.8629	0.8762
FGNN	98,613	0.7880	0.8527
CLNet	154,262	0.8848	0.8943
NCMNet	350,758	1.5419	1.1992
LFGC-DFE	405,717	0.8459	0.8897
KazuNet-Zero	0	0.3811	0.1867
KazuNet (Practical)	45,294	<u>0.2065</u>	<u>0.0432</u>
KazuNet (Strict)	45,294	**0.2050**	<u>0.0432</u>
KazuNet-L (Practical)	239,078	0.2141	**0.0423**

5. Benchmark Results & Theoretical Ablations

5.1. Validation and task-wise results

On held-out validation (Tab. 1), KazuNet (Strict / Practical) reaches mean error 0.2050 / 0.2065 (median 0.0432), cutting error by 74.0% relative to the best external backbone FGNN (0.7880), 59.0% relative to CN-IRLS (0.5039), and 45.8% relative to the zero-parameter learned-free baseline KazuNet-Zero (0.3811), while using fewer parameters (45,294) than every learned external baseline

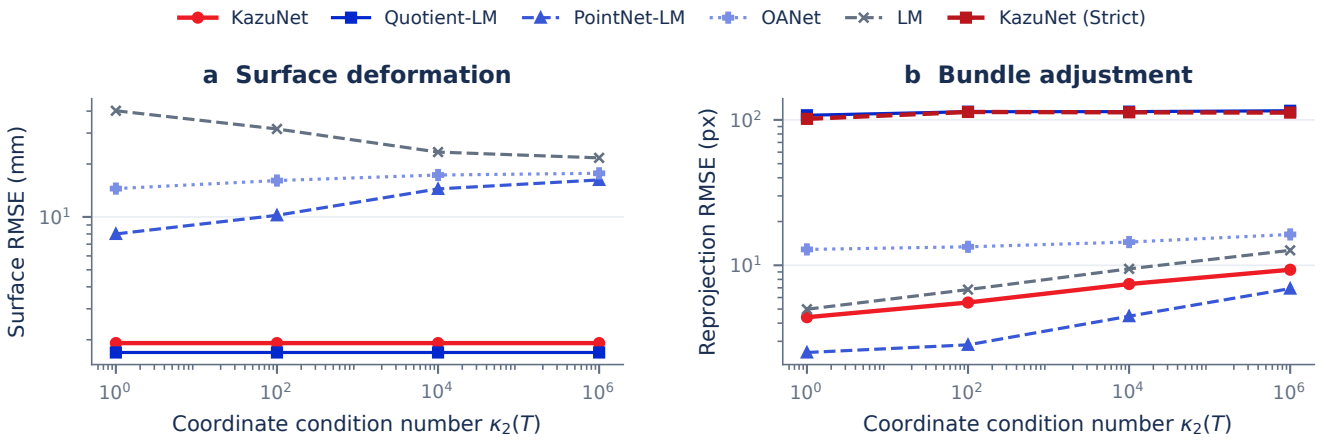


Figure 3. **Coordinate-conditioning ablation at five steps.** Mean errors over the original $C = 30$ gauge grids. (a) Quotient-based methods remain stable under coordinate scaling. (b) Practical spectral truncation improves the bundle-adjustment steps; Strict Mode preserves the full symmetry in Proposition 1.

Table 2. **Mean physical error on eight benchmarks under a five-step budget.** Lower is better; bold and underline mark the best and second-best displayed means. KazuNet (45K) and KazuNet-L (239K) each use one shared checkpoint across tasks. DeepLM uses the author solver with the benchmark residual interface. Complete 19-method results, medians, and subgroups appear in Supp. Sec. C.

Method Cases (unit)	Raw-BAL 140 (px)	Raw-g2o 160 (m)	HPatches 231 (px)	UW-BAL 150 (px)	Wild-BAL 150 (px)	3D-Scans 180 (mm)	Photo-St. 180 (deg)	MATPOWER 180 (pu)
LM	<u>0.773</u>	**0.205**	16.462	2.893	29.461	31.202	31.909	0.1643
DeepLM	**0.738**	<u>0.222</u>	15.902	6.766	33.363	29.286	34.328	0.1820
IRLS-Cauchy	0.903	1.447	<u>5.614</u>	**2.386**	18.700	1.555	10.773	0.0542
GNC-GM	0.883	1.293	**5.398**	<u>2.469</u>	20.168	1.758	17.467	0.1259
HVCE	1.064	4.634	7.493	5.546	56.323	1.873	12.585	**0.0208**
PointNet-LM	0.863	2.120	6.910	3.171	<u>3.562</u>	2.490	3.615	0.0347
CN-IRLS	0.934	2.120	7.366	3.471	8.184	2.844	**3.434**	0.0324
OANet	0.970	1.274	5.694	3.035	39.574	9.489	14.047	0.0846
FGNN	0.932	1.799	6.486	16.819	17.599	5.834	13.905	0.0760
CLNet	0.955	1.528	6.131	4.256	29.887	5.608	24.327	0.0946
NCMNet	1.805	2.272	13.057	9.051	46.513	28.711	31.709	0.1716
LFGC-DFE	0.966	1.699	5.831	5.333	34.569	11.746	20.197	0.0882
KazuNet (45K)	0.827	0.495	6.492	3.414	7.749	<u>0.856</u>	3.832	<u>0.0259</u>
KazuNet-L (239K)	0.820	0.490	6.506	2.564	**1.632**	**0.827**	<u>3.479</u>	0.0260

(55.5K–405.7K). Paired validation comparisons appear in Supp. Sec. B.

Tier I Benchmarks (Raw-BAL, Raw-g2o, HPatches-SIFT). On Raw-BAL ($N = 140$), KazuNet-L (0.820 px) and KazuNet (0.827 px) outperform every robust M-estimator (IRLS 0.903, GNC 0.883 px), every in-repo learned solver (PointNet-LM 0.863, CN-IRLS 0.934 px), and every external backbone (FGNN 0.932 px). On unanchored Raw-g2o ($N = 160$), Foster bridge protection keeps KazuNet at 0.495 m (0.490 m for KazuNet-L) versus 1.274–2.272 m for all neural baselines and 1.293–4.634 m for robust M-estimators, while un-reweighted LM achieves 0.205 m. DeepLM attains 0.738 px on Raw-BAL and 0.222 m on Raw-g2o. On HPatches-SIFT ($N = 231$), KazuNet achieves the overall best six-frame mean (7.944 px across all 19 methods) and an illumination median of 0.4397 px, while Strict Mode lowers the full benchmark mean from 6.492 to 6.353 px. Full subgroup breakdowns are in Supp. Sec. C.

Tier II Benchmarks and multi-task recovery. On the corrupted reconstruction tasks (Tab. 2 and Fig. 4), KazuNet improves recovery across distinct physical objectives. On 3D-Scans, KazuNet (0.856 mm) and KazuNet-L (0.827 mm) reduce error by 45%–47% over the strongest classical solver IRLS-Cauchy (1.555 mm) and by 84.7%–85.3% over CLNet (5.608 mm). On extreme-scale Wild-BAL, KazuNet-

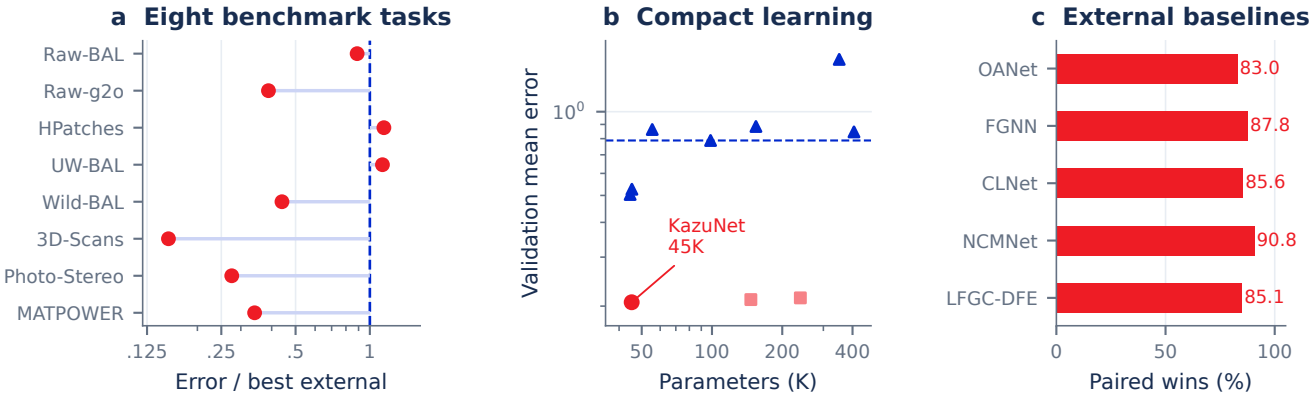


Figure 4. **Learning across tasks.** (a) KazuNet mean error relative to the best of the five external backbones on each task. (b) Validation error versus parameter count; the dashed line is FGNN, the strongest external validation baseline. (c) Paired win rates against the five external backbones, excluding ties. All panels use $T = J = 5$.

L reaches 1.632 px overall first place versus 18.700 px (IRLS), 3.562 px (PointNet-LM), and 17.599 px (FGNN). On Photo-Stereo, KazuNet-L (3.479 deg) and KazuNet (3.832 deg) reduce classical BAR (10.027 deg) and FGNN (13.905 deg) errors by nearly threefold while closely matching the scale-coupled in-repo CN-IRLS (3.434 deg). On UW-BAL, KazuNet-L attains 2.564 px mean and KazuNet lowers median error to 0.917 px (OANet: 2.160 px; IRLS: 1.692 px). On MATPOWER, KazuNet (0.0259 pu) outperforms the five external neural backbones by 2.9×–6.6× and beats HVCE on IEEE-14 (0.0103 vs. 0.0123 pu). Against DeepLM, default KazuNet reduces mean error from 33.363 to 7.749 px on Wild-BAL, 29.286 to 0.856 mm on 3D-Scans, 34.328 to 3.832 deg on Photo-Stereo, and 0.1820 to 0.0259 pu on MATPOWER. Across all 1,371 formal cases (Fig. 4c), 45.3K KazuNet wins 83.0%–90.8% of non-tied paired comparisons against all five external backbones.

Effect of the step budget. Figure 5 tracks mean errors across equal Jacobian budgets $T = J \in \{3, 5, 10, 20\}$. In the budget-ratio comparison (Fig. 6a), five-step KazuNet-family variants surpass the strongest 20-step classical and external-backbone baselines on 3D-Scans by 1.78× and 5.00×, respectively, and on Wild-BAL by 3.11× and 8.38×.

Role of redundancy. KazuNet gains most on overdetermined or corrupted problems. Clean, sparse Raw-g2o SE(2) graphs favor unweighted LM, while three-frame HPatches favors GNC (2.803 vs. 5.053 px); with six frames, KazuNet achieves the lowest mean error.

5.2. Controlled Ablations (A1–A3) & Strict vs. Practical

To isolate the contribution of Stages 3–4 beyond geometric normalization alone, we also evaluate an ablation-only parameter-free control, **Quotient-LM**, which applies Moré–Nielsen damped LM directly in the calibrated quotient coordinates, $\bar{d} = -(1+\lambda)^{-1}\bar{g}_{\text{cal}}$, accepted on $\sum_k \pi_k F_k$ (and using Jacobi-PCG for $n > 256$). On held-out validation ($N = 256$), KazuNet reduces Quotient-LM mean error by 50.7% (0.2065 vs. 0.4192; 164 wins / 0 ties / 92 losses, $p = 1.13\text{e-}12$, Supp. Sec. B).

Tab. 3 and Fig. 3 report all five controlled ablation metrics (A1, A2a, A2b, A2c, A3) isolating Thms. 1–2, Prop. 1, and sparse scaling. On **A1: Scale-1e6**, Quotient-LM (0.0611 m), KazuNet-Zero (0.0627 m), and KazuNet (0.0630 m) all reach the noise floor (2×–20× below HVCE, PointNet-LM, and LM). On **A2: Gauge-GL**, Quotient-LM (1.693 mm on $C = 30$ grids; 1.679 mm on A2a) and KazuNet / KazuNet-L (1.914 / 1.538 mm on $C = 30$ grids, $p = 9.31\text{e-}10$; 1.720 / 1.487 mm on A2a) remain invariant across $\kappa_2(T) \in [1, 10^6]$, where Strict and Practical Modes agree to $< 10^{-3}$. Crucially, on gauge-corrupted bundle adjustment (**A2b Gauge-BAL**) and dimensionless conditioning (**A2c Gauge-Dim**), Practical Mode with learned curvature cuts Quotient-LM error from 65.75 px to 4.366 px (−93.4%) and from 0.415 to 0.084 (−79.8%), confirming the importance of equilibrated spectral truncation alongside learned shrinkage. On **A3: Sparse-Scale** ($n \leq 2,436$), KazuNet / KazuNet-M / KazuNet-L (3.524 / 3.092 / 3.374 mm) cut Quotient-LM error (7.498 mm) by 53.0%–58.8% ($p = 5.5\text{e-}23$). Strict Mode matches Practical Mode identically on 3D-Scans, Photo-Stereo, MATPOWER, A1, and A3 (Supp. Sec. C).

5.3. Implementation, Overhead, and Numerical Equivalence

KazuNet requires only 45,294 `float64` parameters (353.9 KB) and 4.35M neural MACs per step at $K = 64$, scaling as $O(K^2 d)$ independent of state dimension n and residual DOF M, with zero extra Jacobian evaluations ($\Delta J = 0$) during dual-merit backtracking. For $n > 256$,

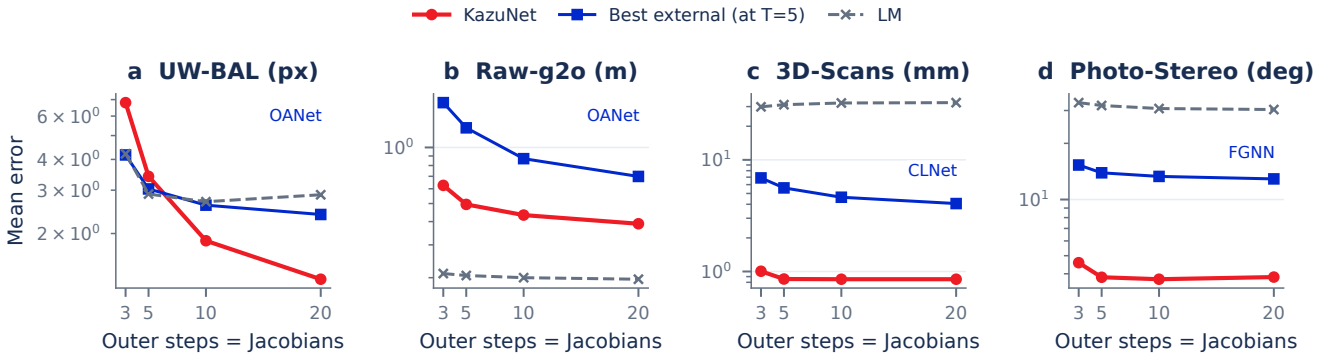


Figure 5. **Error versus outer-step budget.** KazuNet, LM, and the task-specific best external backbone at $T = 5$, held fixed across the four recorded budgets $T = J \in \{3, 5, 10, 20\}$. Each panel names its external comparator; all methods use the same initialization.

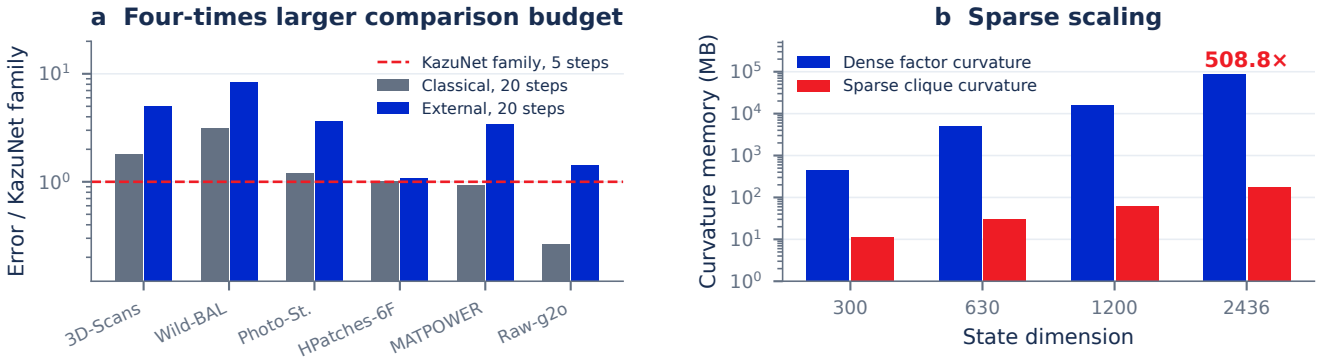


Figure 6. **Budget and memory efficiency.** (a) Best classical and external-backbone errors at 20 steps relative to the best KazuNet-family result at five steps. Values above one favor the five-step result. (b) Sparse and dense factor-curvature memory at batch size 20. The largest case uses $508.8\times$ less memory.

Table 3. **Controlled ablations (A1–A3, $T = J = 5$).** A1: scale-spread SLAM (m); A2a/b/c: gauge-invariant ARAP (mm), BAL (px), and dimensionless stress; A3: sparse meshes (mm). Full 19-solver breakdown is in Supp. Sec. C.

Method	A1	A2a	A2b	A2c	A3
HVCE	0.124	9.891	13.59	0.381	**1.459**
PointNet-LM	0.497	9.874	**3.105**	0.307	7.858
Quotient-LM	**0.061**	1.679	65.75	0.415	7.498
KazuNet-Zero	0.063	1.780	4.865	0.095	7.753
KazuNet (Strict)	0.063	1.720	64.93	0.401	3.524
KazuNet (Pract.)	0.063	1.720	4.366	0.084	3.524
KazuNet-L (Pract.)	0.065	**1.487**	4.370	**0.079**	3.374

sparse clique curvature storage and matrix-free Jacobi-PCG eliminate dense $O(n^3)$ eigensolves and $n \times n$ allocations, reducing factor-curvature tensor memory by $508.8\times$ at batch size 20 and $n = 2,436$ (174.42 MB vs. 88.74 GB, Fig. 6b). On A100/L4 GPUs, canonical `gesvd` and isolated multi-worker CUDA MPS execution preserve deterministic trajectory equivalence across 7,532 five-step runs (relative difference $< 10^{-10}$), whereas two-band approximate SVD induces a $+16.94\%$ error regression on rank-borderline Wild-BAL systems (Supp. Sec. D).

6. Conclusion

We presented KazuNet, a four-stage quotient-geometry optimizer that calibrates factor scales, constructs observable whitened coordinates, predicts Weyl–Procesi invariant curvature coefficients, and selects finite steps via dual-merit projection without extra Jacobians. On 1,371 real and standard benchmark problems at a five-step budget, one compact shared policy outperforms five adapted external neural baselines in 83.0%–90.8% of paired comparisons and reduces error on heavy-outlier 3D scans, wild bundle adjustment, and power-grid state estimation while shrinking sparse curvature memory by up to $508.8\times$. Overall, decoupling geometric normalization from learned anisotropic preconditioning enables one compact shared solver to transfer stably across heterogeneous physical optimization problems under strict low-budget evaluation.

LLM Usage Declaration

The authors take full responsibility for all content: Google Gemini provided manuscript feedback; AI tools generated author-reviewed text and related work, assisted with code, experiments, and data, and autonomously prepared figures, with no use for research ideas, hypotheses, claims, or conclusions.

References

[1] P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. 1, 2

[2] Sameer Agarwal, Noah Snavely, Steven M. Seitz, and Richard Szeliski. Bundle adjustment in the large. In *European Conference on Computer Vision (ECCV)*, pages 29–42, 2010. 1, 5

[3] Sameer Agarwal, Keir Mierle, and The Ceres Solver Team. Ceres solver, 2023. http://ceres-solver.org. 1, 5

[4] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016. 2

[5] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5173–5182, 2017. 5

[6] Jonathan T. Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4331–4339, 2019. 1, 5

[7] Michele Benzi. Preconditioning techniques for large linear systems: A survey. *Journal of Computational Physics*, 182 (2):418–477, 2002. 2, 4

[8] Andrew Blake and Andrew Zisserman. *Visual Reconstruction*. MIT Press, 1987. 1, 5

[9] Luca Carlone, Roberto Tron, Kostas Daniilidis, and Frank Dellaert. Initialization techniques for 3d SLAM: A survey on rotation estimation and its use in pose graph optimization. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2015. 1

[10] Luca Carlone et al. Pose graph SLAM benchmark datasets (intel, manhattan, parking_garage, sphere2500), 2015. OpenSLAM / GTSAM Suite. 5

[11] Tony F. Chan, Gene H. Golub, and Randall J. LeVeque. Algorithms for computing the sample variance. *The American Statistician*, 1983. 1, 3

[12] Brian Curless and Marc Levoy. Stanford 3d scanning repository and AIM@SHAPE mesh collection, 1996. armadillo, stanford-bunny, igea, max-planck, nefertiti, rocker-arm. 5

[13] James Demmel, Laura Grigori, Mark Hoemmen, and Julien Langou. Communication-optimal parallel and sequential QR and LU factorizations. *SIAM Journal on Scientific Computing*, 34(1), 2012. 2, 4

[14] Wolfgang Förstner and Bernhard P. Wrobel. *Photogrammetric Computer Vision*. Springer, 2016. 1, 3

[15] Stuart Geman and Donald E. McClure. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, 52(4):5–21, 1987. 1, 5

[16] Giorgio Grisetti, Rainer Kümmerle, Cyrill Stachniss, and Wolfram Burgard. A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4):31–43, 2010. 1

[17] Paul W. Holland and Roy E. Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics — Theory and Methods*, 6(9):813–827, 1977. 1, 5

[18] Je Hyeong Hong, Christopher Zach, Andrew Fitzgibbon, and Roberto Cipolla. Variable projection methods in bundle adjustment. *SIAM Journal on Scientific Computing*, 2016. 1, 5

[19] Jingwei Huang, Shan Huang, and Mingwei Sun. DeepLM: Large-scale nonlinear least squares on deep learning frameworks using stochastic domain decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10308–10317, 2021. 2, 5

[20] Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. 1

[21] Cornelius Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*, 45(4), 1950. 2, 4

[22] Xin Liu and Jufeng Yang. Progressive neighbor consistency mining for correspondence pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9527–9537, 2023. 2, 5

[23] Jorge J. Moré. The Levenberg–Marquardt algorithm: Implementation and theory. In *Numerical Analysis*, pages 105–116. Springer, 1978. 1, 5

[24] Michael J. D. Powell. A hybrid method for nonlinear equations. In *Numerical Methods for Nonlinear Algebraic Equations*. Gordon and Breach, 1970. 1, 5

[25] Claudio Procesi. The invariant theory of $n \times n$ matrices. *Advances in Mathematics*, 19(3):306–381, 1976. 2

[26] René Ranftl and Vladlen Koltun. Deep fundamental matrix estimation. In *European Conference on Computer Vision (ECCV)*, pages 284–299, 2018. 2, 5

[27] Y. Saidi et al. Variance component estimation and Helmert's method in geodesy. *Journal of Geodesy*, 2003. 1, 3, 5

[28] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Eurographics Symposium on Geometry Processing (SGP)*, pages 109–116, 2007. 1

[29] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of Ill-Posed Problems*. Winston & Sons, 1977. 4

[30] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment — a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms (ICCV)*, pages 298–372, 1999. 1

[31] Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 2

[32] Hermann Weyl. *The Classical Groups: Their Invariants and Representations*. Princeton University Press, 1946. 2

[33] Heng Yang, Pasquale Antonante, Vasileios Tzoumas, and Luca Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. *IEEE Robotics and Automation Letters*, 5(2):1127–1134, 2020. 1, 5

[34] Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2666–2674, 2018. 2, 5

[35] Christopher Zach. Robust bundle adjustment revisited. In *European Conference on Computer Vision (ECCV)*, pages 772–787, 2014. 1, 5

[36] Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao. Learning two-view correspondences and geometry using order-aware network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5845–5854, 2019. 2, 5

[37] Zhen Zhang, Fan Wu, and Wee Sun Lee. Factor graph neural network. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 10720–10730, 2020. 2, 5

[38] Chen Zhao, Yixiao Ge, Feng Zhu, Rui Zhao, Hongsheng Li, and Mathieu Salzmann. Progressive correspondence pruning by consensus learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6464–6473, 2021. 2, 5

[39] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1): 12–19, 2011. 1, 5